Erschienen in: Klosa, Annette (Hrsg.): elexiko. Erfahrungsberichte aus der lexikographischen Praxis eines Internetwörterbuchs. - Tübingen: Narr, 2011. S. 157-172. (Studien zur Deutschen Sprache 55)

Annette Klosa

Von Abbildung bis Wortelement: Weitere Ergänzungen und Änderungen in *elexiko*

elexiko ist ein im Aufbau befindliches Online-Wörterbuch, d. h. es ist ständigen Änderungen in Form von Korrekturen oder Ergänzungen unterworfen.[1] Diese betreffen sowohl die Stichwortliste als auch die lexikografischen Angaben. In diesem Beitrag sollen einige kleinere konzeptionelle Entscheidungen und offene Fragen, die in den anderen Beiträgen in diesem Sammelband noch nicht thematisiert wurden, zusammengefasst werden.

Neuerungen, aber auch noch offene Punkte bestehen bezüglich der Frequenzangaben (vgl. Abschnitt 1.1), der Ausspracheangaben (vgl. Abschnitt 1.2), der Illustrationen (vgl. Abschnitt 1.3) und der Angaben zu Wortbildungsprodukten (vgl. Abschnitt 1.5) in *elexiko*. Andere Angaben wie die Referenzbereichsangabe, die Teil der semantischen Beschreibung ist und auch zur Unterscheidung der Lesarten herangezogen wird, sind zwar in den Wortartikeln erfasst worden (vgl. Abschnitt 1.4), sollen aber noch metalexikografisch ausgewertet werden.

Von den Fragen zur Stichwortliste wurde neben der Frage der Lemmatisierung von Eigennamen[2] auch die Lemmatisierung von Pronomen, Artikeln und adjektivischen Sonderformen geklärt (vgl. Abschnitt 2.1). Offen ist dagegen nach wie vor, ob bzw. wie Wortverbindungen und Wortelemente in *elexiko* lemmatisiert werden sollen (vgl. Abschnitt 2.2).

1. Ergänzungen und Änderungen bei den lexikografischen Angaben

1.1 Frequenzangaben

Im Zuge der Ermittlung der Stichwortkandidaten wurde bei Projektbeginn auch deren Frequenz (und zwar die Summe aller Vorkommen der jeweiligen Grundform und die Summe aller dieser Grundform zugeordneten Flexionsformen) in DEREKO, dem Deutschen Referenzkorpus des IDS, festgestellt.

[1] Vgl. hierzu Klosa (i. Vorb.) und Hahn et al. (2008).

[2] Vgl. hierzu den Beitrag „Die lexikografische Behandlung von Eigennamen in *elexiko*“ von Annette Klosa und Sabine Schoolaert in diesem Band.

Auf diesen Frequenzangaben basierte die Auswahl der endgültigen Stichwörter in *elexiko* (Kandidaten mit einer Frequenz unter 8 wurden nicht in die Stichwortliste aufgenommen; vgl. Schnörch 2005, S. 76).

Die absolute Frequenz eines Stichwortes wird momentan allerdings nicht online angezeigt, weil mittlerweile das *elexiko*-Korpus, ein virtuelles Korpus aus DEREKO, als lexikografische Grundlage dient. Frequenzangaben müssten nun aus diesem Korpus gewonnen werden. Das *elexiko*-Korpus selbst ist zwischenzeitlich wiederum stark angewachsen.[3] Dies bedeutet, dass die ursprünglich ermittelten, absoluten Frequenzangaben heute keine Gültigkeit mehr haben; eine Frequenzangabe kann bei einem dynamischen Monitorkorpus wie dem *elexiko*-Korpus immer nur eine Momentaufnahme sein. Es ist daher nicht sinnvoll, in den Wortartikeln die absolute Frequenz anzugeben.

Auf der Basis der ursprünglich ermittelten Frequenzen wurden für *elexiko* auch Frequenzschichten definiert (vgl. Tab. 1), die vor allem dazu dienen, die Stichwörter sinnvoll zu gruppieren und darauf basierend in Bearbeitungsteilwortschätze einzuteilen.[4] Bei diesen Frequenzschichten handelt es sich also nicht um Häufigkeitsklassen, wie sie etwa das Projekt „Wortschatz-Portal“ definiert: „In (natürlichen) Sprachen kommen die einzelnen Wörter in unterschiedlicher Häufigkeit vor und zwar so, dass relativ wenige Wörter sehr häufig und sehr viele Wörter sehr selten vorkommen. Diesen Umstand spiegeln die Häufigkeitsklassen wider.“ In Relation zum häufigsten Wort wird etwa das Wort *normal* in die Häufigkeitsklasse 10 „(d. h. *der* ist ca. 2^{10} mal häufiger als das gesuchte Wort)“ (http://wortschatz.uni-leipzig.de/help.html) eingeordnet.

Die für *elexiko* definierten Frequenzschichten wurden wiederum in einzelne Gruppen zusammengefasst: niedrigfrequente, durchschnittlich frequente, hochfrequente und höchstfrequente Stichwörter. Die niedrigfrequenten Stichwörter machen etwa 92% der Gesamtmenge (ca. 300.000 Stichwörter) aus, die hoch- und höchstfrequenten Stichwörter zusammen etwa 1%, die durchschnittlich frequenten Stichwörter etwa 7%. Mit ca. 21.000 Stichwörtern entspricht die Menge der durchschnittlich frequenten Stichwörter in etwa einem kleineren einbändigen Bedeutungswörterbuch.

---

[3] Zum *elexiko*-Korpus vgl. Abschnitt 2 der Einleitung in diesem Band.

[4] Zu den Bearbeitungsteilwortschätzen in *elexiko* vgl. Abschnitt 3 der Einleitung in diesem Band. In absehbarer Zeit müsste die Zugehörigkeit der Stichwörter zu den einzelnen Frequenzschichten im *elexiko*-Korpus überprüft werden, da dieses anders zusammengesetzt ist als das Deutsche Referenzkorpus (DEREKO) des IDS.

Frequenz-schicht	Stichwortfrequenz-summen	Gruppen	Teilwortschätze zur Bearbeitung
I	0	niedrig-frequente Stichwörter	Stichwörter mit automatisch generierten Angaben[5]
II	1-9		
III	10-50		
IV	51-100		
V	101-500		
VI	501-1.000	durchschnittlich frequente Stichwörter	
VII	1.001-5.000		
VIII	5.001-10.000		
IX	10.001-50.000	hochfrequente Stichwörter	„Lexikon zum öffentlichen Sprachgebrauch“[6]
X	50.001-100.000		
XI	100.001-500.000		
XII	500.001-1.000.000	höchstfrequente Stichwörter	
XIII	1.000.001-8.900.000		

Tab. 1: Frequenzschichten und -gruppen in *elexiko*

Eine Angabe zur Zugehörigkeit zu einer der Frequenzschichten erfolgt derzeit bei solchen Wörtern, die noch nicht redaktionell bearbeitet sind, zusammen mit anderen automatisch ermittelten Angaben (vgl. Abb. 1).[7]

glücklich

Dieses Stichwort gehört im *elexiko*-Korpus der Frequenzschicht IX (10.001-50.000 mal belegt) an. Es ist in 18 verschiedenen Zeitungen oder Zeitschriften aus 26 Jahrgängen belegt.

Abb. 1: Angaben zur Frequenzschicht und der Korpusbelegung im Wortartikel *glücklich*

Neben der Zuordnung zu einer Frequenzschicht wird außerdem angegeben, in wie vielen verschiedenen Quellen und aus wie vielen verschiedenen Jahrgängen das Wort im *elexiko*-Korpus belegt ist (Stand 2009), damit die Nutzer einen Eindruck davon bekommen können, wie verbreitet das Wort ist. Da sich das Korpus auch weiterhin verändern wird, müssen diese Werte nach einer gewissen Zeit im Korpus überprüft und gegebenenfalls in den Wortartikeln aktualisiert werden.

[5] Zu Stichwörtern mit automatisch ermittelten Angaben vgl. Abschnitt 3 der Einleitung in diesem Band.

[6] Zum *Lexikon zum öffentlichen Sprachgebrauch* vgl. Abschnitt 3 der Einleitung in diesem Band.

[7] Vgl. hierzu auch Abbildung 1 in Abschnitt 3 der Einleitung in diesem Band.

Bei redaktionell bearbeiteten Wörtern wird auf diese Angabe verzichtet, da fast alle Wörter, die derzeit bearbeitet sind bzw. werden, zwischen 10.000 und 500.000 mal im Korpus belegt sind, und also zu den hochfrequenten Stichwörtern zählen. Durch die Zugehörigkeit eines Stichwortes zum *Lexikon zum öffentlichen Sprachgebrauch* wird die Einordnung des entsprechenden Stichwortes in die Frequenzschichten IX bis XI also impliziert, weshalb die Frequenzschichten online nicht extra angezeigt werden.

Zukünftig wäre vorstellbar, nach dem Vorbild der Angaben zur zeitlichen Verteilung der Gebrauchshäufigkeiten der Stichwörter im Neologismenwörterbuch (2005ff.) die Frequenzangaben auch in *elexiko* mit ihrer zeitlichen Entwicklung in Diagrammform zu gestalten.[8] Auch hierbei würde auf die Angabe absoluter Frequenzen verzichtet, weil deren Aussagekraft bei hochgradig polysemen Stichwörtern eher eingeschränkt ist. Aus den Angaben zur relativen Wortfrequenz könnten sich die Nutzer aber einen Eindruck von der Erstbelegung und der Verteilung im *elexiko*-Korpus verschaffen. Da das *elexiko*-Korpus sich zukünftig weniger vom Umfang her als in der Zusammensetzung verändern wird, könnten aus einer Darstellung der Entwicklung der relativen Frequenzen über einen bestimmten Zeitraum hinweg unter Umständen auch Schwankungen deutlich werden.

1.2 Ausspracheangaben

Von Anfang an wurde in der Struktur der *elexiko*-Wortartikel eine lesartenbezogene Ausspracheangabe vorgesehen, durch die lesartenbezogene Aussprachevarianten berücksichtigt werden können (z. B. *Tenór* ‘Sänger in der höchsten Stimmlage’ versus *Ténor* ‘Grundgehalt einer Äußerung’). Allerdings wurde während der Konzeptionsphase von *elexiko* noch nicht entschieden, ob die Aussprache mithilfe von Lautschriftumschreibungen, Tondateien oder beidem erfolgen soll. Ebenso war offengeblieben, ob das Stichwort in isolierter Aussprache oder in Aussprache im Satzzusammenhang gezeigt werden soll.

Eine Sichtung der Praxis in anderen Online-Wörterbüchern[9] ergab, dass hier sowohl phonetische Umschrift als auch Hörbeispiele als auch die Kombination beider Alternativen vorkommen, wobei eine Kombination aus phonetischer Umschrift und Tondateien grundsätzlich am sinnvollsten erscheint, weil nicht alle Nutzer immer die Möglichkeit haben dürften, eine Audiodatei

[8] Solche Diagramme, so genannte „Verlaufsstatistiken“, finden sich beispielsweise auch im *DWDS* – dem *Digitalen Wörterbuch der deutschen Sprache des 20. Jahrhunderts*.

[9] Ausgewertet wurden: *American Heritage Dictionary* bei Dictionary.com, *Cambridge Dictionaries Online*, *DWDS*, *LEO*, *Merriam-Webster Online* und *PONS – Das Sprachenportal*.

abspielen oder rezipieren zu können. Sind Audiodateien vorhanden, verdeutlichen diese häufig das Stichwort in isolierter Aussprache. Meist werden solche Ausspracheangaben gezielt hergestellt, sind also nicht das Ergebnis natürlicher Sprachproduktion. Die isolierte Aussprache eines Wortes kann für Lerner des Deutschen sinnvoller sein, weil das Phänomen der Koartikulation, welches die Aussprache verwischen oder undeutlich machen kann, nicht auftritt. Andererseits widerspricht die Angabe einer isolierten, gezielt hergestellten Aussprache eines Stichworts dem sonst in *elexiko* geltenden Prinzip, dass alle Angaben auf natürlichsprachlichen Äußerungen beruhen (vgl. Haß 2005b, S. 7). Für eine natürlichsprachliche Angabe der Aussprache kann diese nur aus Tonausschnitten aus gesprochensprachlichen Korpora bestehen. Außerdem würde es dem Ansatz von *elexiko*, Vielfältigkeit und Varianz der Sprache zu beschreiben, widersprechen, wenn nur eine isolierte, gezielt hergestellte Ausspracheangabe zu einem Stichwort gegeben würde.

Vor diesem Hintergrund wurde mit der Auswahl von Hörbelegen für redaktionell bearbeitete Stichwörter aus der „Datenbank Gesprochenes Deutsch“ im „Archiv für Gesprochenes Deutsch (2009)“ des IDS begonnen. Dabei werden ein bis drei Hörbelege ausgewählt, die das Stichwort möglichst in Hochlautung, gleichmäßiger, natürlicher Sprechweise und guter Tonqualität enthalten sollen. Der gewählte Tonausschnitt isoliert das Stichwort nicht, sondern bietet 5 Sekunden Kontext vor und 5 Sekunden Kontext nach dem Stichwort, sodass die jeweilige Lesart kontextuell aufscheinen kann. Vorerst werden diese Hörbelege allerdings nicht lesartenbezogen angeordnet werden können, da die Gesamtmenge an qualitativ guten Hörbelegen dafür nicht ausreicht.

Online werden die Hörbelege in einem separaten Bildschirmfenster angezeigt (vgl. Abb. 2), das durch Klicken auf ein Lautsprechersymbol neben der Lemmazeichengestaltangabe in der normalen Artikelansicht zu öffnen ist. Die einzelnen Hörbelege können durch Auswahl eines Hörformats neben dem entsprechenden Transkriptausschnitt geöffnet werden, parallel kann der entsprechende Transkriptausschnitt gelesen werden. Über einen Link lassen sich weitere Hörbelege in der „Datenbank Gesprochenes Deutsch“ suchen.

Für noch nicht bearbeitete Stichwörter ist in einem weiteren Schritt an die Anzeige automatisch ausgewählter Hörbelege gedacht, soweit das Stichwort überhaupt in den gesprochensprachlichen Korpora belegt ist. Eine begleitende Erarbeitung von Ausspracheangaben in phonetischer Umschrift (IPA) für alle Stichwörter wäre grundsätzlich wünschenswert, ist aus praktischen Gründen im Projekt aber derzeit nicht zu leisten.

Hörbelege für *Bier*

Die folgenden Belege wurden manuell ausgewählt.
WMA / MP3 führt zu Tonausschnitten im gleichnamigen Format.

WMA / MP3	… die werden natürlich sofort erst mal bewirtet und zwar mit einem ordentlichen Schnaps oder mit eim **Bier** und Brot nich wahr während die Hebamme schon an ihrem Werk ist und es ist selbstverständlich …
WMA / MP3	… besondere Problematik des Begriffes Subkultur bei uns sichtbar die Frage wie man in Mannheim ein **Bier** bestellt erscheint zunächst relativ sinnlos aber sie bekommt einen relativen Sinn sobald wir …

[Zur Startseite der „Datenbank Gesprochenes Deutsch"]

Abb. 2: Präsentation von Hörbelegen zum Wortartikel *Bier*

1.3 Illustrationen

Einige Stichwörter im *elexiko*-Demonstrationswortschatz[10] enthalten auf der Grundlage der Vorüberlegungen zu Illustrationen für *elexiko* in Müller-Spitzer (2005) Fotos, welche die Bedeutungserläuterung begleiten (z. B. in den Wortartikeln *Bahn* [Lesart 'Zug'], *Kathedrale* [Lesart 'Bischofskirche'], *Rollstuhl* [Lesart 'Gefährt']). Jeweils ein oder mehrere Foto(s) können durch Klicken auf eine Schaltfläche „Abbildung(en)" geöffnet werden und erscheinen dann mit der Legende „Beispiel(e) für [Stichwort]" in einem separaten Bildschirmfenster (vgl. Abb. 3).

Nach Abschluss der Arbeit am Demonstrationswortschatz wurden zunächst keine neuen Abbildungen in *elexiko* integriert, weil eine Überprüfung der lexikografischen Praxis in internationalen Online-Wörterbüchern (eingefordert in Müller-Spitzer 2005, S. 224) noch ausstand, die aber für ein gründliches Illustrationskonzept unabdingbar erschien. Zwischenzeitlich wurde diese Analyse in Reinhard (2007) vorgelegt, woraus sich einige Vorschläge für die Illustrationspraxis in *elexiko* und Online-Wörterbüchern allgemein ergeben haben, die allesamt unter den obersten Prinzipien der Benutzerfreundlichkeit und Übersichtlichkeit stehen:[11]

– Dichte und Auswahl: Bei allen Lemmata, die einer visuellen Ergänzung bedürfen, sollten Illustrationen erscheinen, da es online keine Platzbeschränkung gibt. Zu rein dekorativen Zwecken sind Abbildungen dagegen abzulehnen.

[10] Zum *elexiko*-Demonstrationswortschatz vgl. Abschnitt 3 der Einleitung in diesem Band.

[11] Vgl. hierzu ausführlicher Reinhard (2007, S. 255ff.).

– Platzierung: Die Illustrationen sollten in unmittelbarer Nähe der entsprechenden Bedeutungserläuterung stehen, am besten in einem vergrößerbaren, zusätzlichen Bildschirmfenster, sodass Bedeutungserläuterung und Illustration gemeinsam rezipiert werden können. Auf keinen Fall sollten Illustrationen lesartenübergreifend präsentiert werden.
– Größe und Format: Am besten sind Abbildungen in einem einheitlichen Layout, die nur so groß wie nötig angezeigt werden sollten.
– Gestaltung: Am besten sind Zeichnungen, weil die zu illustrierenden Gegenstände hierin in abstrahierter und typisierter Form präsentiert werden können. Aus Kostengründen sind auch Fotografien akzeptabel.
– Legenden sollten möglichst immer erscheinen, da eine Illustration für sich semantisch offen und unendlich deutbar ist.
– Ein Illustrationsindex ist ratsam.
– Illustrationstypen: Besonders unikale (d. h. nur einen Gegenstand zeigende) und strukturelle (d. h. einen Gegenstand in Beziehung zu einer größeren Struktur zeigende) Illustrationen sind empfehlenswert; daneben auch aufzählende (d. h. mehrere Beispiele für einen Gegenstand präsentierende) und nomenklatorische (d. h. einen Fachwortschatz darstellende) Illustrationen, wenn von diesen auf die entsprechenden Wortartikel verlinkt wird.[12]

Kathedrale Lesart 'Bischofskirche'

Mit **Kathedrale** wird (vor allem mit Bezug auf Frankreich, Spanien, England und die Schweiz) die meist mittelalterliche, große Kirche eines Bischofssitzes bezeichnet.

[Abbildung(en)]

Abb. 3: Bedeutungserläuterung zu *Kathedrale*, Lesart 'Bischofskirche' mit zugehörigen Illustrationen

[12] Zu diesen und anderen Typen von Illustrationen vgl. Hupka (1989a und b).

Auf der Grundlage dieser Überlegungen, aber auch aufgrund praktischer Gegebenheiten (s. u.) wird die Illustrationspraxis im *elexiko*-Demonstrationswortschatz nun auf das *Lexikon zum öffentlichen Sprachgebrauch* ausgeweitet. Wenn das Stichwort sinnvollerweise einer visuelle Ergänzung bedarf, sollen mehrere Abbildungen (vorzugsweise Fotos) in einem möglichst einheitlichen Layout ergänzt werden. Dabei wird unikalen oder strukturellen Abbildungen der Vorzug gegeben. Zunächst war allerdings zu überprüfen, welche redaktionell bearbeiteten Stichwörter überhaupt illustriert werden können:[13] über 250 Nomen, circa 75 Verben, über 50 Adjektive und 20 Adverbien wären demzufolge derzeit Kandidaten für Illustrationen (Stand 2010). Weil in *elexiko* Illustrationen weiterhin nur parallel zu Bedeutungserläuterungen und lesartenbezogen angeboten werden sollen, werden nicht bearbeitete Stichwörter dagegen auch zukünftig nicht illustriert werden.

Daneben wurde überprüft, welche Möglichkeiten es zur kostenlosen Gewinnung von Illustrationen gibt. Hierbei sind verschiedene Online-Datenbanken mit Zeichnungen, Fotografien, Comics und Videos ausgewertet worden (z. B. pixelio.de, Wikimedia Commons). Das Ergebnis ermutigt, was die Menge kostenlos zur Verfügung stehender Illustrationen (vor allem Fotos) betrifft, wirft aber auch einige neue Probleme auf: So sind die Fotos hinsichtlich ihrer Art, Qualität, Auflösung und Größe sehr unterschiedlich. Ebenso unterschiedlich sind die Vorgaben dazu, wie die Quellenangabe erfolgen muss. Wichtig bei der Integration in die *elexiko*-Wortartikel ist auch, dass ein weiterer Download der Abbildungen verhindert werden müsste, weil das die Nutzungsbedingungen der Bilddatenbanken vorschreiben können.

Nach der Klärung dieser eher technischen Fragen sollen die schon bearbeiteten Stichwörter im *Lexikon zum öffentlichen Sprachgebrauch* illustriert werden, daneben wird Bildmaterial für die noch zu bearbeitenden Wortartikel gesammelt und den Lexikografen zur Auswahl vorgeschlagen. Ziel ist, das ganze *Lexikon zum öffentlichen Sprachgebrauch* mit möglichst vielen, qualitativ hochwertigen Illustrationen zu versehen, wobei diese zunächst aus praktischen Gründen eher auf unikale Abbildungen beschränkt bleiben müssen. Wünschenswert ist auch der Ausbau um strukturelle Illustrationen, z. B. für Stichwörter wie *Kopf* (Lesart ‘Teil des Körpers’), *Haupt* (Lesart ‘Kopf’), *Arm* (Lesart ‘Körperteil’), *Bein* (Lesart ‘Körperteil’), *Fuß* (Lesart ‘Teil des Beins’) als Teile von *Körper* (Lesart ‘Gestalt’) oder *Auge* (Lesart ‘Sehorgan’), *Mund*

[13] Eine lesartenbezogene Überprüfung steht noch aus; die Zahl der möglichen Illustrationen wird sich vermutlich noch erhöhen, wenn nicht einzelne Stichwörter, sondern einzelne Lesarten gezählt werden.

(Lesart ‘Teil des Gesichts’), *Ohr* (Lesart ‘Hörorgan’), *Nase* (Lesart ‘Sinnesorgan’), *Gesicht* (Lesart ‘Vorderseite des Kopfes’) als Teile von *Kopf* (Lesart ‘Teil des Körpers’).[14] Mithilfe der Beschriftung der einzelnen Teile des gezeigten Ganzen und einer Verlinkung dieser Bezeichnungen auf die entsprechenden Wortartikel würden die Wortartikel nicht nur über die Illustration stärker vernetzt werden, sondern es würde den Nutzern auch eine onomasiologische Zugriffsmöglichkeit angeboten.

1.4 Referenzbereichsangabe

Die Referenzbereichsangabe, die von Anfang an in der Wortartikelstruktur von *elexiko* vorgesehen war, soll u. a. dazu dienen, zwischen Lesarten eines Lexems zu disambiguieren, wenn diese auf unterschiedliche, größtenteils ontologische Bereiche verweisen (vgl. hierzu Haß 2005a, S. 172). Eine erste projektinterne Prüfung hatte ergeben, dass bestehende Ontologien als Kategorieninventar für die lexikografische Arbeit zu begrenzt sind.

Diese Angaben wurden und werden daher derzeit relativ frei und eher experimentell ausgefüllt. So ist im Wortartikel *Bord* für die Lesart ‘Rand’ der Referenzbereich „Schiffsteil“, für die Lesart ‘Brett’ der Referenzbereich „Möbelstück“ und für die Lesart ‘Böschung’ der Referenzbereich „Landschaftsteil“ eingetragen. Für bestimmte Stichwörter (z. B. Abstrakta) kann nur schwer ein Referenzbereich angegeben werden, und die Klassifizierung von Eigenschaftsprädikatoren ist beispielsweise (noch) sehr grob. Diese Angabe wird daher online vorerst nicht angezeigt, obwohl sie großes Potenzial bietet. Wird etwa bei Stichwörtern wie *Universität*, *Fachhochschule*, *Schule*, *Gymnasium*, *Realschule* in den Lesarten ‘Bildungseinrichtung’ usw. als Referenzbereich „Bildungseinrichtung“ erfasst, könnte diese Angabe zukünftig dazu dienen, onomasiologische Zugriffe auf die Wortartikel zu ermöglichen. Die redaktionell bearbeiteten Wortartikel sollen deshalb unter diesem Gesichtspunkt im Rahmen des Projektes BZV*elexiko*[15] geprüft werden. Außerdem soll hier eingehender untersucht werden, ob – und wenn ja, in welcher Weise – bestehende Ontologien doch als Grundlage für ein festeres Kategorieninventar für die Referenzbereichsangaben dienen können.

1.5 Angaben zu Wortbildungsprodukten

Die Ermittlung von Wortbildungsprodukten und ihre Erfassung in den Wortartikeln war zwar in der Anfangsphase des *elexiko*-Projektes konzipiert wor-

[14] Die genannten Stichwörter sind alle redaktionell bearbeitet.

[15] Zum Projekt BZV*elexiko* vgl. das Vorwort in diesem Band.

den,[16] musste dann aber aus praktischen Gründen zunächst zurückgestellt werden (vgl. Haß 2005b, S. 12). Geplant ist, im Angabebereich „Wortbildungsproduktivität" solche Wörter aus der *elexiko*-Stichwortliste einzutragen (und mit den entsprechenden Stichwörtern zu verlinken), zu denen das Stichwort selbst die Ableitungs- oder Kürzungsbasis ist, oder in denen das Stichwort als Teil einer Zusammensetzung auftritt. Die Angaben sollen dabei im besten Fall lesartenbezogen erfolgen, weil zu verschiedenen Lesarten eines Stichworts unterschiedliche Wortbildungsprodukte vorliegen können (vgl. van der Colff 1998 und Holly 1986).

In der Zwischenzeit konnte die Arbeit an diesem Angabebereich im Rahmen des Projektes BZV*elexiko* aufgenommen werden. Ziel dieser Arbeiten ist es, zu möglichst vielen Stichwörtern (zunächst aber vor allem zu einfachen, d. h. nicht gebildeten Wörtern) Wortbildungsprodukte wie Komposita und Derivate in der *elexiko*-Stichwortliste automatisch zu ermitteln und in geeigneter, ohne redaktionellen Eingriff realisierbarer Form darzustellen. Die Wortbildungsprodukte könnten beispielsweise sortiert nach Wortbildungsarten angezeigt werden (zum Stichwort *Computer* etwa die Derivate *computerisieren* oder *Computerei* und Komposita wie *Computerfachmann* oder *Bordcomputer*) oder nach Frequenz der Wortbildungsprodukte im *elexiko*-Korpus.[17]

Besonders bei lexikografisch noch nicht bearbeiteten Stichwörtern können Nutzer auf diese Weise einen Eindruck von den durch Wortbildung entstehenden Vernetzungen im Wortschatz bekommen.[18] Die Angaben zu Wortbildungsprodukten in *elexiko* sollen daneben auch neuartige Formen des Zugriffs eröffnen.

2. Fragen der Lemmatisierung

2.1 Die Lemmatisierung von Pronomen, Artikeln und adjektivischen Sonderformen

Im Zuge der praktischen Artikelarbeit wurde deutlich, dass für bestimmte Wortgruppen die Frage des Stichwortansatzes (vgl. hierzu generell Schnörch 2005) in der ursprünglichen Konzeption offengeblieben war oder den tatsächlichen Anforderungen nicht entsprach. Hiervon waren insbesondere Artikel und Pronomen und eine Reihe von Adjektiven betroffen.

---

[16] Vgl. hierzu genauer Klosa (2005, S. 151ff.).

[17] Zu weiteren Einzelheiten vgl. die Internetseiten des Projektes BZV*elexiko* (www.ids-mannheim.de/lexik/BZVelexiko).

[18] Warum Angaben zur Wortbildungsproduktivität sonst noch sinnvoll sind, beschreiben z. B. Barz (1995) und Bergenholtz (2000).

Da für ein Online-Wörterbuch generell nicht die Notwendigkeit besteht, Platz einsparen zu müssen, wurde entschieden, nach dem Genus unterscheidende Pronomen und Artikel einzeln zu lemmatisieren. Bei den Artikeln und Pronomen werden also jeweils alle Formen des Nominativ Singular als einzelne Lemmata angesetzt (z. B. *der*, *die*, *das*; *dein*, *deiner*, *deine*, *deines*). Zugleich ermöglicht das Medium Internet, dass diese Artikel über Hyperlinks so verbunden werden, dass den Nutzern der Zusammenhang zwischen den einzelnen Wörtern bewusst wird. Derzeit geschieht dies mithilfe der Bedeutungserläuterung, in der auf die Formen in den anderen Genera hingewiesen wird (vgl. Abb. 4); zugleich sind diese Formen mit den entsprechenden Wortartikeln verlinkt.

Zusätzlich wird auf den entsprechenden Eintrag in *grammis*, dem grammatischen Informationssystem des IDS, verlinkt, wo weitere morphologische, syntaktische und semantische Informationen aufgerufen werden können. In den grammatischen Angaben zu diesen Pronomen und Artikeln wird in *elexiko* das gesamte Flexionsparadigma im jeweiligen Genus abgebildet.

deine Lesart ‘Pronomen‘

deine ist ein feminines Possessivpronomen (maskulin: **deiner**, neutral: **deines** bzw. **deins**).

> Mama, da lag eine Tasche hinter dem Auto, ist das **deine**?“ (Rhein-Zeitung, 05.11.1996, Mamas Tasche.)

Weitere Informationen:
Zu morphologischen, syntaktischen und semantischen Informationen vgl. den Eintrag Possessiv-Pronomen in grammis, dem grammatischen Informationssystem des IDS.

Abb. 4: Bedeutungserläuterung im Wortartikel *deine*, Lesart ‘Pronomen’

Adjektive vom Typ *innere*, *innerer*, *inneres* oder *linke*, *linker*, *linkes* oder *äußerste*, *äußerster*, *äußerstes* werden anders behandelt: Sie werden unter einer Lemmazeichengestaltangabe zusammengefasst, die ohne die Endungen erscheint, also z. B. *inner-*, *link-*, *äußerst-*. Damit wird die Tatsache berücksichtigt, dass die Formen aller Genera sich in ihrer Bedeutung und Verwendung nicht unterscheiden. Bei diesen Autosemantika liegt der Fokus im Wortartikel auf den semantischen Angaben. Bei den Artikelwörtern und Pronomen hingegen, die als grammatische Wörter in *elexiko* der Klasse der Synsemantika zugeordnet werden (vgl. Haß 2005a, S. 170), liegt der Schwerpunkt insbesondere auf ihrer grammatischen Beschreibung, sodass hier die getrennte Lemmatisierung vorzuziehen ist.

Der Vergleich zwischen den Stichwortstrecken *Deichwache – Deixel* und *linieren – Linksabbieger* in Tabelle 2 verdeutlicht die unterschiedlichen Lemmatisierungsprinzipien, mit denen für beide Gruppen im Rahmen des allgemeinen Konzeptes Ausnahmeregelungen gefunden wurden, die die jeweiligen Eigenheiten und Anforderungen berücksichtigen.

Deichwache – Deixel	**linieren – Linksabbieger**
Deichwache	linieren
Deichwesen	liniert
deiktisch	Linierung
dein	link
deindustrialisiert	Link
Deindustrialisierung	**link-**
deine	Linke
deiner	linken
deines	Linker
deinesteils	linkerseits
deinethalben	linkisch
deinetwegen	Linkohr
deinetwillen	Linkrusta
Deismus	links
Deist	links außen
deistisch	links orientiert
Deixel	Linksabbieger

Tab. 2: Unterschiedliche Lemmatisierung von Pronomen und adjektivischen Sonderformen in *elexiko*

2.2 Die Lemmatisierung von Wortverbindungen und Wortelementen

Ursprünglich wurde bei der Planung von *elexiko* davon ausgegangen, dass dieses als umfassendes Informationssystem zur deutschen Gegenwartssprache sowohl Einwortlemmata wie Stichwörter zu Mehrwortverbindungen und Wortbildungsmitteln enthalten sollte (vgl. Haß 2005b, S. 12). Die Beschreibung von usuellen Wortverbindungen und Mehrwortlemmata sollte innerhalb eines eigenen Moduls in *elexiko*, das damals noch als Portal fungierte,[19] geschehen

[19] Vgl. Klosa (2008, S. 3).

(vgl. ebd, S. 16). Deshalb wurden und werden z. B. in den *elexiko*-Wortartikeln feste Wortverbindungen wie Phraseologismen (z. B. „der ganz normale Wahnsinn“ zum Stichwort *normal*, Lesart ‘üblich’) oder Redensarten (z. B. „Lieber reich und gesund als arm und krank“ zum Stichwort *gesund*, Lesart ‘wohlauf’) explizit weitgehend ausgeklammert. In den *elexiko*-Wortartikeln werden dagegen nicht idiomatische, feste Verbindungen (z. B. „unter normalen Bedingungen“ zum Stichwort *normal*, Lesart ‘üblich’, oder „organisch völlig gesund sein“ zum Stichwort *gesund*, Lesart ‘wohlauf’) im Angabebereich „Typische Verwendungen“[20] erfasst. Kollokatoren aus binären Verbindungen (z. B. „normaler Unterricht“ zum Stichwort *normal*, Lesart ‘üblich’, oder „gesunde Kinder“ zum Stichwort *gesund*, Lesart ‘wohlauf’) werden in den Wortartikeln schließlich im Angabereich „Semantische Umgebung und lexikalische Mitspieler“[21] dargestellt.

In der Zwischenzeit ist aus diesem ursprünglichen Modul das Projekt „Usuelle Wortverbindungen“ entstanden, das Artikel zu solchen *Festen Wortverbindungen* innerhalb von OWID, dem Online-Wortschatzinformationssystem Deutsch des Instituts für Deutsche Sprache, veröffentlicht und usuelle Wortverbindungen auf den eigenen Internetseiten „Wortverbindungen online“ beschreibt. Vor diesem Hintergrund muss entschieden werden, ob sich *elexiko* auch weiterhin auf die Beschreibung von Einzellexemen (bzw. zukünftig auch von Wortbildungsmitteln, s. u.) beschränken wird, oder ob zumindest die Mehrwortlexeme in *elexiko* aufgenommen werden, die bei der Arbeit mit den Korpusbelegen und -befunden zu einem Einzellexem auftauchen. Dies hätte den Vorteil, dass stärker als bislang feste Wortverbindungen zu in *elexiko* bearbeiteten Stichwörtern auf der gleichen Korpusbasis lexikografisch beschrieben werden könnten. Auf der anderen Seite ist das Projekt mit der Beschreibung der Einzelwörter auf Jahre ausgelastet. Eine grundsätzliche Entscheidung zu dieser Frage sollte aber noch vor Ende der Bearbeitung des *Lexikons zum öffentlichen Sprachgebrauch* getroffen werden.

Die Frage der Behandlung von Wortelementen (d. h. Wortbildungsmitteln) in *elexiko* ist nach wie vor offen (vgl. hierzu Haß 2005b, S. 12): Weder konnte bislang entschieden werden, welche Wortbildungsmittel in *elexiko* als Stichwörter (so genannte Wortelementlemmata) aufgenommen, noch, in welcher Form sie lexikografisch beschrieben werden sollen. Dabei würde *elexiko*

[20] Vgl. hierzu den Beitrag „Die typischen Verwendungen in *elexiko*“ von Christine Möhrs in diesem Band.

[21] Vgl. hierzu den Beitrag „Neue Überlegungen und Erfahrungen zu den lexikalischen Mitspielern“ von Annette Klosa und Petra Storjohann in diesem Band.

mit der lexikografischen Behandlung von Wortbildungsmitteln „sowohl den Forderungen der Wörterbuchforschung[22] nachkommen wie auch eine in vielen gegenwartssprachlichen Wörterbüchern verbreitete Tradition fortsetzen“ (Klosa 2005, S. 154f.).

Aus der Arbeit an der automatischen Analyse der Wortgebildetheit der Stichwörter[23] sind aber Erkenntnisse dazu zu erhoffen, welche Wortbildungsmittel (Affixe, Konfixe) wie häufig für die Bildung der *elexiko*-Stichwörter genutzt werden. Außerdem werden besonders häufige, Reihen bildende Bestandteile von Zusammensetzungen zu erkennen sein. Die Auswahl und Beschreibung der Wortbildungsmittel in *elexiko* kann daher auf diesen Materialien fußen.

Grundsätzliches Ziel bleibt also, dass *elexiko* die Wortbildungsmittel lemmatisieren und beschreiben soll, „wie sie in der Sprache der öffentlichen Diskussion realisiert werden“ (Klosa 2005, S. 157); mit der praktischen Umsetzung kann vermutlich aber erst nach Ende der Bearbeitung des *Lexikons zum öffentlichen Sprachgebrauch* begonnen werden.

3. Literaturverzeichnis

3.1 Wörterbücher

American Heritage Dictionary – bei Dictionary.com. Internet: http://dictionary.reference.com/ (Stand: 30.04.2010).

Cambridge Dictionaries Online. Internet: http://dictionary.cambridge.org (Stand: 30.04.2010).

DWDS – Das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts. Internet: http://www.dwds.de/ (Stand: 30.04.2010).

elexiko (2003ff.). In: Institut für Deutsche Sprache (Hg.): OWID – Online-Wortschatz-Informationssystem Deutsch. Mannheim. Internet: www.elexiko.de (Stand: 30.04.2010).

Feste Wortverbindungen (2007ff.). In: Institut für Deutsche Sprache (Hg.): OWID – Online-Wortschatz-Informationssystem Deutsch. Mannheim. Internet: www.owid.de (Stand: 30.04.2010).

LEO – Web-Angebot mit Online-Wörterbüchern Deutsch-Englisch, Deutsch-Französisch, Deutsch-Spanisch, Deutsch-Italienisch, Deutsch-Chinesisch. Internet: www.leo.org (Stand: 30.04.2010).

Merriam-Webster Online. Internet: www.merriam-webster.com (Stand: 30.04.2010).

---

[22] Vgl. z. B. Barz (2002) oder Schmidt (2000).

[23] Vgl. hierzu Abschnitt 3 der Einleitung in diesem Band.

Neologismenwörterbuch (2005ff.). In: Institut für Deutsche Sprache (Hg.): OWID – Online-Wortschatz-Informationssystem Deutsch. Mannheim. Internet: www.owid.de (Stand: 30.04.2010).

PONS – Das Sprachenportal. Internet: http://de.pons.eu/ (Stand: 30.04.2010).

3.2 Forschungsliteratur

Barz, Irmhild (1995): Komposita im Großwörterbuch Deutsch als Fremdsprache. In: Pohl, Inge/Ehrhardt, Horst (Hg.): Wort und Wortschatz. Beiträge zur Lexikologie. Tübingen, S. 13-24.

Barz, Irmhild (2002): Die Wortbildungsmittel im de Gruyter Wörterbuch Deutsch als Fremdsprache. In: Wiegand, Herbert Ernst (Hg.): Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des „de Gruyter Wörterbuchs Deutsch als Fremdsprache". (= Lexicographica. Series Maior 110). Tübingen, S. 105-121.

Bergenholtz, Henning (2000): Lexikographie und Wortbildungsforschung. In: Barz, Irmhild et al. (Hg.): Praxis- und Integrationsfelder der Wortbildungsforschung. (= Sprache – Literatur und Geschichte 18). Heidelberg, S. 19-30.

Colff, Ari van der (1998): Die Komposita in Langenscheidts Großwörterbuch Deutsch als Fremdsprache. In: Wiegand, Herbert Ernst (Hg.): Perspektiven der pädagogischen Lexikographie des Deutschen. (= Lexicographica. Series Maior 86). Tübingen, S. 193-207.

Hahn, Marion/Klosa, Annette/Müller-Spitzer, Carolin/Schnörch, Ulrich/Storjohann, Petra (2008): *elexiko* – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem. Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche. In: Klosa (Hg.), S. 57-85. Internet: www.ids-mannheim.de/pub/laufend/opal/privat/pdf/opal08-1_hahn-klosa-mueller-spitzer.pdf (Stand: 18.05.2010).

Haß, Ulrike (2005a): Das Bedeutungsspektrum. In: Haß (Hg.), S. 163-181.

Haß, Ulrike (2005b): *elexiko* – Das Projekt. In: Haß (Hg.), S. 1-17.

Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York.

Holly, Werner (1986): Wortbildung und Wörterbuch. In: Lexicographica 2/1986, S. 195-213.

Hupka, Werner (1989a): Wort und Bild. Die Illustrationen in Wörterbüchern und Enzyklopädien. (= Lexicographica. Series Maior 22). Tübingen.

Hupka, Werner (1989b): Die Bebilderung und sonstige Form der Veranschaulichung im allgemeinen einsprachigen Wörterbuch. In: Hausmann, Franz Josef et al. (Hg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie. 1. Teilbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK) 5.1). Berlin/New York, S. 704-726.

Klosa, Annette (2005): Wortbildung. In: Haß (Hg.), S. 141-162.

Klosa, Annette (2008): Vorwort. In: Klosa (Hg.), S. 3-4. Internet: www.ids-mannheim.de/pub/laufend/opal/privat/pdf/opal08-1_vorw.pdf (Stand: 11.05.2010).

Klosa, Annette (Hg.) (2008): Lexikografische Portale im Internet. (= OPAL – Online publizierte Arbeiten zur Linguistik 1/2008). Mannheim. Internet: www.ids-mannheim.de/pub/laufend/opal/privat/opal08-1.html (Stand: 30.04.2010).

Klosa, Annette (i. Vorb.): The lexicographical process II: online dicitonaries. In: Gouws, Rufus H. et al. (Hg.): Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography. Berlin/New York.

Müller-Spitzer, Carolin (2005): Vorüberlegungen zu Illustrationen in *elexiko*. In: Haß (Hg.), S. 204-226.

Reinhard, Christina-Doreen (2007): Untersuchungen zu Illustrationen in Online-Wörterbüchern. Unveröffentlichte Magisterarbeit an der Ruprecht-Karls-Universität Heidelberg, Abteilung Germanistische Sprachwissenschaft.

Schmidt, Rosemarie (2000): Grammatik und Lexikographie. Wortbildungsmittel im zweisprachigen Wörterbuch anhand deutscher, schwedischer und russischer Beispiele. In: Bayer, Josef/Römer, Christine (Hg.): Von der Philologie zur Grammatiktheorie. Peter Suchland zum 65. Geburtstag. Tübingen, S. 303-313.

Schnörch, Ulrich (2005): Die *elexiko*-Stichwortliste. In: Haß (Hg.), S. 71-90.

3.3 Internetressourcen

Archiv für Gesprochenes Deutsch. Internet: http://agd.ids-mannheim.de/html/index.shtml (Stand: 30.04.2010).

BZV*elexiko* – Benutzeradaptive Zugänge und Vernetzungen in *elexiko*. Internet: www.ids-mannheim.de/lexik/BZVelexiko (Stand: 30.04.2010).

Datenbank Gesprochenes Deutsch (DGD). Internet: http://dsav-wiss.ids-mannheim.de/DSAv/DSAVINFO.HTM (Stand: 30.04.2010).

DEREKO – Das deutsche Referenzkorpus. Internet: www.ids-mannheim.de/kl/projekte/korpora (Stand: 30.04.2010).

grammis – das grammatische Informationssystem des Instituts für Deutsche Sprache. Internet: http://hypermedia.ids-mannheim.de/index.html (Stand: 30.04.2010).

pixelio.de – Deine kostenlose Bilddatenbank für lizenzfreie Fotos. Internet: www.pixelio.de (Stand: 30.04.2010).

Wikimedia Commons – a database of 6,237,097 freely usable media files. Internet: http://commons.wikimedia.org/wiki/Main_Page (Stand: 30.04.2010).

Wortschatz-Portal der Universität Leipzig. Internet: http://wortschatz.uni-leipzig.de (Stand: 30.04.2010).

Wortverbindungen online – Plattform des Projekts Usuelle Wortverbindungen. Internet: http://wvonline.ids-mannheim.de (Stand: 30.04.2010).